

Polysemy Effects and Chronological Memory

Francesco Fumarola

(Dated: December 6, 2016)

The existence of a polysemy effect in episodic memory is demonstrated through an analysis of data from the experiments of Lohnas et al. (2015) and Healey and Kahana (2016). Three word-length related features are reported: (1) the average distance between the serial positions of consecutively recalled words is an increasing function of the second word’s length; (2) the recall of words in the order in which they were presented is more likely to occur for shorter words; (3) longer words are easier to recall than shorter words from the same list. These phenomena are reproduced by modeling the fact that shorter words have a larger number of meanings (polysemy). This is shown to be at the root of our tendency to recall verbal sequences in forward order, and a conjecture is put forth regarding the general methods we employ to store information on the chronology of events.

I. EXPERIMENTAL OBSERVATIONS

Free-recall experiments are one of the key tools for the controlled investigation of episodic memory. In a typical free-recall experiment, a list of words is presented to a subject who is then asked to recall them in any order. A key variable is the difference between the positions of two consecutively recalled words, called a “lag”. E.g., if the 5th word in the list is remembered right after the 8th, the corresponding lag is $\tau = -3$ (Kahana, 2012).

A large database of free-recall data has been recently assembled for the University of Pennsylvania Electrophysiology of Encoding and Retrieval Study. Collapsing the data described in Lohnas et al. (2015) with those described in Healey and Kahana (2016) yields a total of 12,016 free-recall trials, all performed with lists of 16 words. The wordpool from which the lists were assembled contains 1638 words of up to 6 syllables.

My analysis of this database has revealed three word-length related features, displayed in Figure 1:

(1) The average lag through which a word is recalled is an increasing function of its length;

(2) The probability $p(\tau, \lambda)$ that, if a word of length λ is retrieved, it is retrieved with a lag τ depends strongly on λ only for $\tau = +1$ (“sequential” recall);

(3) The average recall probability of words from a given list is an increasing function of their length (this fact was first reported by Katkov et al., 2014).

Observations (1) and (2) entail that $p(1, \lambda)$ is a decreasing function of λ , as previously reported in (Fumarola, 2016).

Over the last two decades, experimental results have pushed psychological research toward the idea that word length may not affect free recall directly, but as a statistical indicator of other word properties (Neath et al., 2003; Campoy, 2008; Jalbert et al., 2011).

Recent fMRI measurements (Musz and Thompson-Schill, 2015) have shown that the typical neural response to the presentation of a word exhibits a strong statistical dependence on the word’s “polysemy” – the variability of the word’s meaning with context. If the polysemy of a word affects importantly its perception at the neural level, one would suppose that it may also affect our ability to memorize it and recall it. In a suspicious coincidence, word length happens to be a strong statistical indicator of polysemy; more specifically, a word’s length is known to correlate negatively with its number of meanings. This has been proven in a host of languages ranging from French to Maori (Zipf, 1949; Guiter, 1974; Sambor, 1984; Rother, 1994).

The aim of this paper is to show that the phenomena listed above are interdependent components of a single “polysemy effect”, born out of a mechanism through which the semantic variability of input enforces our chronological memory. In the next section, I will begin by defining a simple mathematical model of polysemy. I will then use it to simulate the various stages of a standard free-recall experiment. All the experimental observations I mentioned above will be reproduced and justified by the model. At last, some general conclusions will be drawn on the role of chronology in memory-storage processes.

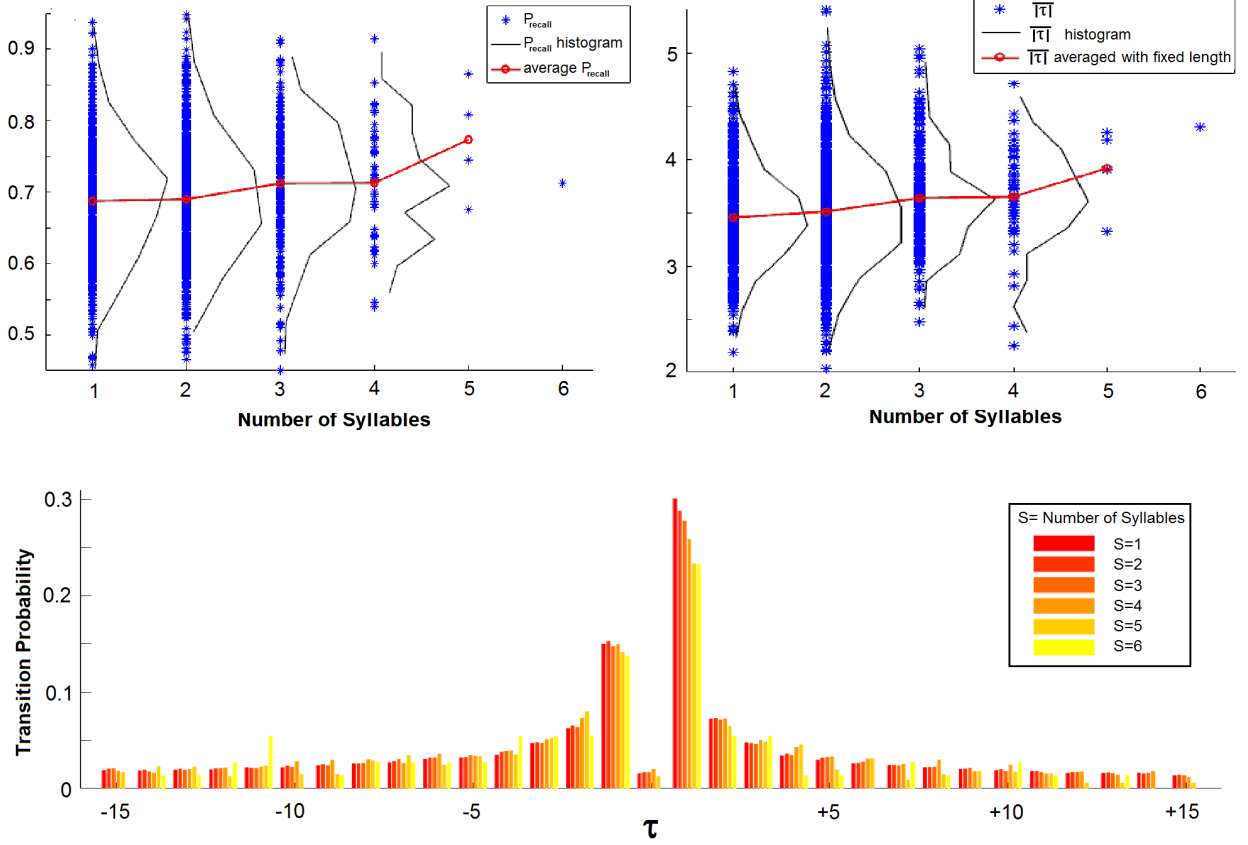


FIG. 1: Word-length dependent phenomena in the experiments of Lohnas et al. (2015) and Healey and Kahana (2016). Word lengths used in the experiments ranged between 1 and 6 syllables, but only one six-syllable word was included. Upper-left panel: total recall probability as a function of word length; blue dots correspond to individual words, black lines are histograms, and averages over all words with the same length are shown in red; the correlation coefficient is 0.11 ($p < 10^{-4}$). Upper-right panel: averaged absolute value of the serial-position lag, both for individual words (blue dots, with black lines as histograms) and averaged over all word with the same length (red curve); the correlation coefficient is 0.14 ($p < 10^{-5}$). Lower panel: average probability that a word, if recalled, is recalled with a certain lag, plotted as a function of the lag and color-coded for the number of syllables in the word. A striking monotonous behavior emerges for the sole lag $\tau = +1$ (“sequential” transitions); data for individual words (not shown) yield a correlation coefficient -0.17 ($p < 10^{-5}$) between sequential-transition probability and word length.

II. A TOY MODEL OF POLYSEMY

A graph representation of neural attractors has recently allowed for a quantitative understanding of some key features of associative recall (Romani et al., 2013), and would therefore be a natural approach to our problem. I will focus here on the general properties that the system of attractors must satisfy in order to reproduce the experimental results, postponing to future work the identification of networks with the attractor structure I will demonstrate.

This task is facilitated by the knowledge garnered over the last two decades in the field of verbal recall. As was first proven through retrieved-context theories (Howard and Kahana, 2002), the object of verbal retrieval in episodic memory is never the word itself, but rather the “context” corresponding to its presentation. The attractor graph, therefore, has to include both a set of verbal nodes (the various meanings of each word) and a set of nonverbal nodes that provide the contextual “bath” in which verbal meanings are immersed. We may call the latter a “contextual reservoir”.

I will consider here the case where the reservoir is a circular graph, and each word is linked to two contextual states. A word can then only be monosemous or bisemous: a monosemous word has a single meaning linked to two “contexts” of the reservoir, while a bisemous word is linked to two contexts through two different meanings. These two types of word can be interpreted as corresponding to different word-lengths (or syllable counts) with the bisemous words being shorter.

(This is a simplified variant of the lattice model of (Fumarola, 2016). Some of the results I will obtain replicate results from the lattice version.)

Suppose the vocabulary contains N short words and N long words. Since each long word is represented by a single node and each short word by 2 nodes, there are $3N$ verbal nodes, connected to the reservoir through $4N$ links. If every site of the reservoir is connected to one verbal node, the full graph (including both the reservoir and the verbal nodes) contains $7N$ nodes. Observables must be predicted by averaging over all inequivalent graph configurations corresponding to the same value of N , of which there are $\frac{1}{4} [(2N+1)(N-1)! + (2N)!/(2^N N)]$. Figure 1 displays one of the configurations for $N = 3$.

All thought process is represented as a simple random walk on the graph, where the graph’s edges are undirected and have all the same weight. Once a word has been presented, the system will ‘recognize’ it only when it chances upon any of the verbal nodes associated

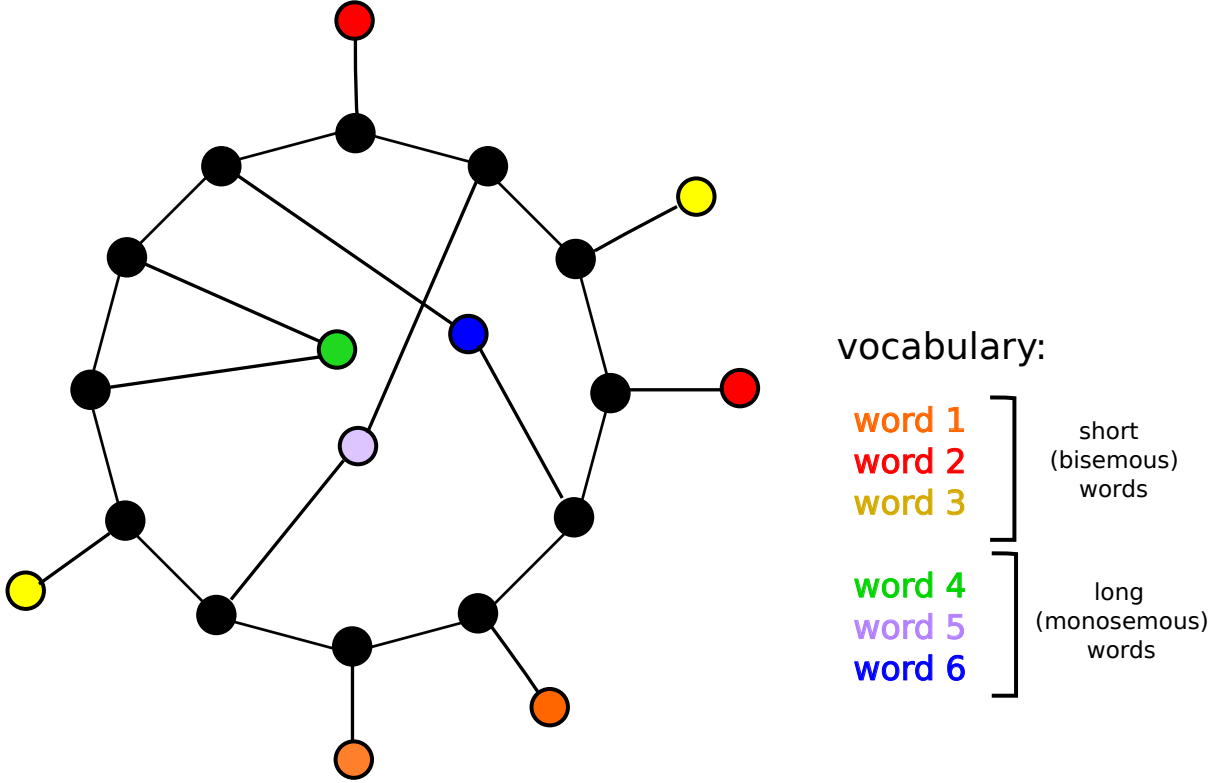


FIG. 2: A configuration of the model with a 6-word vocabulary. The contextual reservoir consists here of 12 states, represented as black circles and joined into a ring. The vocabulary includes 3 bisemous words and 3 monosemous words, listed in colors to the right. The meanings of each word are represented by circles filled with the word's color. The whole graph has 21 nodes, 12 non-verbal nodes and 9 verbal ones. Thought processes are represented as random walks on this graph, thus allowing for a full simulation of free-recall experiments.

to it. A memory trace is conserved in all nodes visited recently, where “recently” means within a time lapse longer than the duration of a typical free-recall experiment.

To simulate free-recall experiments, I submit word-lists that are permutations of the full vocabulary. By discounting permutations of words of the same type within the vocabulary, it is sufficient to check $(2N)!/N!^2$ inequivalent lists. The initial location of the system is a random site in the circular reservoir; each time a new word is presented, the process random-walks until it recognizes the word, and only then the next word is presented. Thus verbal memories are formed.

Experimentally, the memory test begins after a delay of several seconds from the end of the presentation stage. Accordingly, I reset the position of the system at random before

starting a new random walk whose purpose is to retrieve memories. This amounts to ignoring recency effects (Murdock, 1962).

Memories are retrieved in the order in which they are encountered. When no new memory is encountered for a maximal time T , the retrieval process comes to an end. The maximal inter-response time T and the half-size N of the vocabulary are the only free parameters in the model.

III. COMPARISON WITH DATA

The upper left panel of Fig. 3 displays a typical example of the distribution of transition probabilities, computed numerically for $N = 6$ and $T = 1000$. Its most conspicuous feature is the emergence, for short words, of one of the universal features of free-recall data (Kahana, 2012): the “sequential peak”, i.e., a sharp peak in correspondence of the lag $\tau = +1$. This is the same peak we noticed in the distribution in the lower panel of Fig. 1.

I mentioned that only for $\tau = +1$ the transition probability exhibits a strong dependence on length, and this is readily seen from the lower panel of Fig. 1. In the upper-left panel of Fig. 3, we find that the same is true in the simulations (with the sole difference that there are now two word lengths).

The probabilities of transitions between words during the recall process approach steady asymptotes for $T \rightarrow \infty$. In the upper-right panel of Fig. 3, I show the asymptotic values of $p(+1)$ (i.e. of the sequential transition probability) as functions of the number N of short and long words in the vocabulary. While both curves decrease, the short-word curve lies always above the long-word curve. This, too, replicates the experiments – the lower panel of Fig. 1 – where the probability that a word is recalled sequentially decreases with its length.

The lower left panel of Fig. 3 shows the values of the average recall probabilities for short and long words, plotted as functions of T , for two representative values of N . Increasing N while keeping T fixed decreases both probabilities, but their comparative magnitudes remain in the same order, i.e., long words are systematically easier to recall. This is just another of the observations we made from experimental data, discernible in the upper left panel of Fig. 1.

The lower right panel of Fig. 3 shows the absolute value $|\tau|$ of the lag, averaged over transitions in which short or long words are recalled. $\overline{|\tau|}$ tends to a steady asymptote for

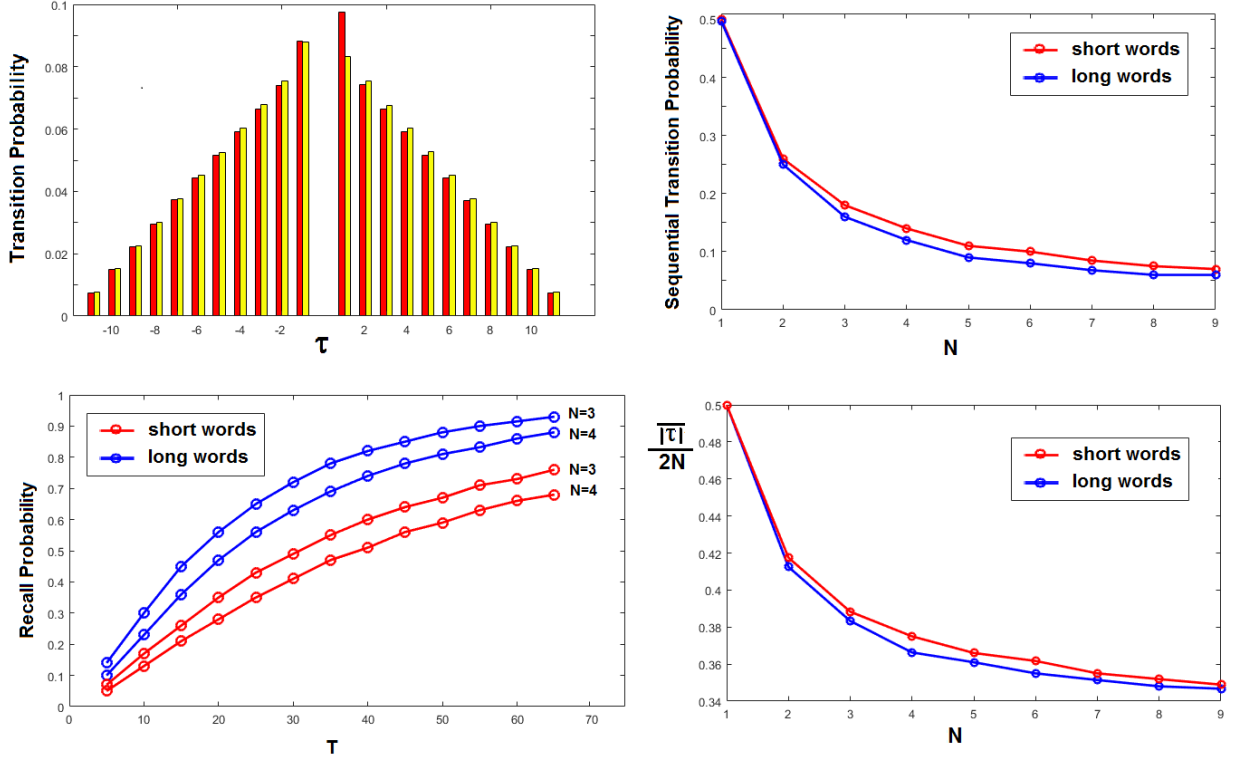


FIG. 3: Data from simulations of the model of Fig. 2. Upper left panel: a typical distribution of transition probabilities, computed for $N = 6$ and $T = 1000$. Upper right panel: asymptotic value (for large T) of the probability of sequential transitions to short and long words during the recall process, averaged over all words of the same length and plotted as a function of N . Lower left panel: recall probabilities of short and long words for two representative sizes of vocabulary, plotted as functions of T . Lower right panel: average lag for transitions to short/long words, normalized by the length of the lists and plotted as a function of the half-size of the vocabulary.

large T , and this is the value plotted in the figure (normalized by the number of words in the lists, which is $2N$). The mean lag for short words is always longer, precisely as we had found in the upper right panel of Fig. 1.

We may conclude that, in spite of the model's simplicity, its agreement with the experimental data is substantial.

We can also grasp intuitively, in terms of polysemy, why sequential transitions are favored for short words. Since each short word we present has multiple meanings, the system is likely to give it the meaning closest at hand, which depends on the system's current position –

mostly, on the last word presented. A short-word memory tends therefore to cluster with the word-memory that precedes it; this enhances the sequential recall of a short word, whether it is preceded by a long word or by another short word.

The enhancement of backward contiguous transitions is less relevant for two reasons: first, because a word only clusters with the word that follows if that word is a short word, which makes the effect at $\tau = -1$ smaller than the effect at $\tau = +1$; second, because the backward contiguous recall of short and long words are enhanced to the same extent whenever they are followed by a short word, i.e., the effect acts equally on words of different length.

IV. DISCUSSION OF RESULTS

A. Hermeneutic irreversibility

The first conceptual result of the above is a new way of understanding the observed asymmetry of recall data with respect to the direction of time during the presentation process, that is, the asymmetry of the distribution in the lower panel of Fig. 1 around $\tau = 0$.

Suppose that all words were monosemous. The retrieval process would then be independent on the system's history during presentation, and would only depend on the structure of the list. The distribution of transition probabilities could then be calculated from first principles and would be $p(\tau) = \frac{M-|\tau|}{M(M-1)}$, where M is the length of the lists. This is completely invariant under the reversal operation $\tau \rightarrow -\tau$.

Therefore, if we used an ideal language in which all words have a single meaning, our verbal memory would be invariant under the inversion of the time arrow, and the plot in the lower panel of Fig. 1 would be perfectly symmetric. By introducing some degree of polysemy, we force the receiver of verbal input to choose one of many possible understandings, and this process is intrinsically irreversible.

Semantic ambiguities are thus found to play the same role that the multiplicity of microscopic configurations plays in thermodynamics. A thermodynamic system can behave irreversibly because each of its macroscopic states can correspond to multiple microscopic configurations. In our model, words function as “macroscopic” observables, meanings as “microscopically” defined states.

The psychological interpretation of this finding is clear. During a purely receptive task the intensity of time’s flow is proportional to our hermeneutic freedom. The relationship between polysemy and irreversibility may be understood as implementing a principle of least effort (Zipf, 1949): we don’t need to remember the chronology where it does not play a role in determining the meaning of events.

B. The sequential peak

In experiments, the asymmetry parameter $\eta = P(\text{forward}) - P(\text{backward})$ has the value $\eta = 0.1$ if contiguous transitions ($\tau = \pm 1$) are counted in, and $\bar{\eta} = -0.05$ if they are excluded. The same qualitative behavior is found in simulations, e.g. for $N = 3$ we find $\eta = 0.01$ and $\bar{\eta} = -0.003$. Hence, ignoring contiguous transitions is enough to eliminate the forward asymmetry completely, both in simulations and experiments, leaving in fact a slight degree of backward asymmetry.

Thus, information on the chronology is stored almost entirely in the sequential peak (the peak in transition probabilities corresponding to $\tau = +1$).

All conscious observation of a polysemous sensorial input may be modeled similarly, and forward asymmetry will arise likewise. We expect therefore that, when we are paying attention to consecutive events, the corresponding memories will be stored in such a way that the chronological order is the most probable order of recall. We will refer to this fact as “sequentiality”.

C. Fragmentation of the lists

Suppose that a logical discourse, e.g. a narrative, is interrupted abruptly and immediately followed by a random word. Instinctively, we will try to fit the meaning of the word into the ongoing narrative.

Our ease in doing so will depend on the particular word. Short words allow for a wider semantic choice, so they are easier to fit into an existing narrative, while long words have fixed meanings, probably difficult to graft coherently into the current narrative. With a long word, therefore, our mind perceives a “change of scenery”, and assumes that a *new* narrative is beginning. (For more on this argument, see Fumarola, 2016).

This is precisely what we found to happen within the model. Since $p(1, \lambda)$ is a decreasing function of λ , episodic memory will break each list into multiple “narratives”, and long words will tend to correspond to the points of fragmentation of the list.

The probability peak at $\tau = 1$ tends to guide the recall process through each narrative from start to finish. Then it surrenders the command back to free association, allowing associative mechanisms to decide which narrative will be recalled next.

D. Retrieving accurate chronologies

It is not completely obvious that sequentiality is the optimal way to preserve chronological information. After all, even where information was stored sequentially, the retrieval process has a finite probability of occurring in non-chronological orders.

Moreover, even without the peak, the system may still be able to assess in what order the events occurred – through some computation on the probability landscape encountered in the region of semantic space where the presentation procedure has taken place.

If the chronological order is the “most” probable, however, the system has a particularly simple way of singling it out with arbitrary accuracy. It is sufficient to re-explore the same contextual area a large number of times, and choose the ordering of memories that has been experienced most often during this re-exploration.

The more strictly “sequential” our memory storage is (i.e. the larger $p(+1)$), the less time we will take to perform the iterative sampling needed to establish a chronology with arbitrary accuracy. This procedure may act as a probabilistic version of the long-term enhancement of the principal eigenvector that plays a similar role in neural networks.

E. Optimal sequentiality

A final question (whose answer I will merely sketch) is what determines the precise height of the sequential peak.

If the sequential peak is too low, the number of iterations needed to find the most probable ordering will become large, and the iterative sampling procedure slow. Devoting more than a fraction of a second to ordering any sequence of past events may be impractical.

If, on the contrary, the sequential peak is too high, all associative retrieval of a memory

will be blocked. That follows from the normalization of probabilities. When a memory can only be accessed chronologically, it is consequently not available for associative reasoning – comparisons, deductions – and becomes useless for most conceivable purposes.

Thus, sequentiality and retrievability are in conflict and a trade-off between the two requirements is necessary. A memory must stay available for associative reasoning, and yet its chronology needs to be trackable through iterative sampling. From these two constraints, the optimal value of the $p(\tau = +1)$ is determined.

This optimization process will depend on the use we make of a given memory. Some memories are meaningless without their chronology, others are just as useful without. We have seen here an example of this. After splitting a word list into separate narratives, our model adheres to the policy that the chronology of events within a narrative is more important than the ordering of the various narratives, and it stores memories accordingly.

If the links in the contextual graph can be created and destroyed over time, the optimization process will be dynamical. The degree of sequentiality of a memory may decrease over time or stay constant. Of course it can be ruled out, statistically, that it will increase back toward its original value. Chronological information obeys the entropic principle, and if we call $T^{I.S.}$ the time that iterative sampling needs to provides the correct chronology for a set of memories, we may write $\frac{dT^{I.S.}}{dt} \geq 0$.

A memory that hasn't yet been overwritten by the time it loses its chronology is, by definition, no longer episodic and de facto transitioning into long-term memory.

V. CONCLUSIONS

The existence of a polysemy effect in free recall has been proven through the analysis of data from the experiments of Lohnas et al. (2015) and Healey and Kahana (2016).

Three word-length features were observed: (1) the average distance between the serial positions of consecutively recalled words is an increasing function of the second word's length; (2) the recall of words in the order in which they were presented is more likely to occur for shorter words; (3) longer words are easier to recall than shorter words from the same list.

The well-documented fact that shorter words have a larger number of meanings allowed to build a simple model for these phenomena. Within the model, long words are easier to recall because their single meaning is more easily accessible during the retrieval process.

Short words are distributed more widely in semantic space, and the meaning we assign to them during presentation is accordingly harder to retrieve.

Our ability to choose among various meanings results also in a higher probability to recall words in the order in which they were presented. This is analogous to the mechanism that makes macroscopic motion irreversible in thermodynamics. Short words, having a larger number of meanings, exhibit this tendency toward sequential recall to a greater degree.

The probability peak at sequential transitions has a crucial cognitive function. It allows to reconstruct the chronology of past events, which would otherwise be lost. This could be done with arbitrary accuracy through a process of iterative sampling, by exploring multiple times the region of semantic space where the memories are stored, and by choosing the ordering that occurs most often. The sequential peak, however, must not be too high or it would hinder associative operations.

Future experimental work may provide a much better understanding of the polysemy effect. For instance, we have argued that the polysemy effect has no direct connection to the phonology or duration of words. Experiments with words that are outliers in the polysemy/word-length scatter plot would provide a definitive test of this assumption.

The concept of orthographic neighborhood (on which experiments have been carried out by Jalbert et al., 2011) grades almost continuously into the notion of polysemy. Words with a large orthographic neighborhood, therefore, may behave just like polysemous words where sequentiality is concerned. This, too, should be the object of further study.

I would like to thank Michael J. Kahana, of the University of Pennsylvania, for generously sharing the data obtained in his laboratory.

VI. BIBLIOGRAPHY

Campoy G. (2008). The effect of word length in short-term memory: Is rehearsal necessary? *Quarterly Journal of Experimental Psychology*, 61:5, 724-734.

Fumarola F. (2016). Verbal perception and the word length effect. Preprint, arXiv:1607.05530.

Guiter H. (1974). Les relations fréquence-longueur-sens des mots (langues Romaines et Anglais), in “Atti del Congresso Internazionale di Linguistica”, 14(4): 373-381. Amsterdam: Benjamins.

- Healey M. K. and Kahana M. J. (2016). A four-component model of age-related memory change. *Psychological Review*, 123(1): 23-69.
- Howard M. W. and Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, 46: 269-299.
- Jalbert A., Neath I., Bireta T. J., and Surprenant A. M. (2011). When does length cause the word length effect? *J. Exp. Psychol. Learn. Mem. Cogn.* 37: 338-353.
- Kahana M. J. (2012). *Foundations of Human Memory*. New York, NY: Oxford University Press.
- Katkov M., Romani S., and Tsodyks M. (2014). Word length effect in free recall of randomly assembled word lists. *Frontiers of Computational Neuroscience* 8:129-133.
- Lohnas L.J., Polyn S.M., and Kahana M.J. (2015). Expanding the scope of memory search: Modeling intralist and intralist effects in free recall, *Psychological Review*, 122(2): 337-363.
- Murdock B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, 64 (2): 482-488.
- Musz E. and Thompson-Schill S.L. (2015). Semantic variability predicts neural variability of object concepts. *Neuropsychologia*. 76: 41-51.
- Neath I., Bireta T.J., and Surprenant A.M., (2003). The time-based word length effect and stimulus set specificity, *Psychonomic Bulletin Rev.* 10(2):430-4.
- Romani S., Pinkoviezky I., Rubin A., and Tsodyks M. (2013). Scaling laws of associative memory retrieval. *Neural Computation* 25:2523-2544.
- Rothe U. (1994). Wortlänge und Bedeutungsmenge: Eine Untersuchung zum Menzerath-schen Gesetz an drei romanischen Sprachen. In: Köhler R., Boy J. (eds.), *Glottometrika* 5, 101-112. Bochum: Brockmeyer.
- Sambor J. (1984). Menzerath's law and the polysemy of words. In *Glottometrika* 6: 94-114. Bochum: Brockmeyer.
- Zipf G. K. (1949). *Human behaviour and the principle of least effort*. Cambridge: Addison-Wesley.